

# DIGITAL IMAGE WATERMARKING RESISTANT TO GEOMETRIC AND REMOVAL ATTACKS IN THE WAVELET TRANSFORM DOMAIN

Ray Naegle

[rnaegl01@shepherd.edu](mailto:rnaegl01@shepherd.edu)

Department of Computer Science, Mathematics, and Engineering  
Shepherd University, Shepherdstown, WV 25443-3210

## ABSTRACT

This paper presents an autocorrelation function (ACF)-based watermarking scheme that is resistant to both removal and geometric attacks. A host image is processed to give itself periodic autocorrelation. A periodic watermark is then embedded in the perceptually significant blocks of a host image's wavelet transform coefficients. The attack is estimated using the resulting autocorrelation peaks of the host image to approximate and reverse a geometric attack. The watermark is then extracted from the restored image. Our extensive experimental results demonstrate that xxx.

**Index Terms**— Image processing, Wavelet transforms, Correlation

## 1. INTRODUCTION

In recent years, the Internet has become a staple of modern life. The increased penetration of broadband network access has made it easier for individuals to share information and communicate with one another. At the same time, this increased freedom to share data poses a threat to copyright holders, whose intellectual property can be shared illegally.

Lee *et al.* [1] describe an ACF-based watermarking scheme in the spatial domain which increases the autocorrelation peaks of a watermarked image. In their method, the cover image is first filtered to isolate the noise which would interfere with the embedded watermark signal.

In another publication, Lee and Lee [2] propose a similar watermarking procedure in the wavelet transform domain. The wavelet transform domain is used by modern image compression algorithms. This allows us to embed different periodic watermark signals with varying strengths, based on the wavelet transform decomposition level and sub-band, which results in increased resistance to removal attacks.

The work described in this paper is based heavily upon the findings of Lee *et al.* [1], [2]. The proposed method combines the strengths of the schemes described. By

increasing the autocorrelation of the cover image before embedding the watermark, the resulting peak strength is higher. A watermark is then applied to the image, using properties of the wavelet transform domain to embed the signal with higher strength than in the spatial domain, while keeping visual distortion to a minimum.

## 2. THE PROPOSED EMBEDDING SCHEME

Two steps are involved in the watermark embedding process. First, we will modify the cover image before embedding to ensure a high autocorrelation after removal attacks. An ACF-based approach is limited by the strength of its autocorrelation peaks—without them, the geometric attack cannot be reversed, and the watermark cannot be extracted. Second, we will embed the watermark in the wavelet domain of the modified cover image to maintain high invisibility and robustness to removal attacks. The details of these two steps are explained in the following subsections.

### 2.1 Cover image preprocessing

The watermark signal, a kind of additive noise, can be isolated from the cover image by taking the difference between the cover image and the Wiener filtered image. Most images contain noise to begin with, so we seek to minimize interference between such noise and our watermark. To do this, the noise is isolated from the cover image by

$$E = I - fI, \quad (1)$$

where  $I$  represents the cover image, and  $fI$  is the Wiener filtered image.

The resulting noise  $E$  is segmented into blocks of size  $M/2 \times M/2$ , and the average block  $r(x, y)$  is calculated by

$$r(x, y) = \frac{\sum_{n=1}^N v_n(x, y)}{N} \quad (3)$$

where  $v_n(x, y)$  represents the segmented blocks of  $E$ . The average block  $r$  is then up-scaled to have size  $M \times M$ , to make the changes made here more resistant to removal attacks.

To increase the autocorrelation of the image, the segmented blocks  $v_n$  are treated as vectors. For each  $v_n$ , the up-scaled average block  $R$  is treated as a vector, and modified to have the same length as each  $v_n$ , as shown by

$$R_n = \frac{v_n \cdot R}{\|R\|} \quad (4)$$

The resulting reference vector  $R_n$  is used to calculate the difference vector  $d_n$  for each  $v_n$  as

$$d_n = R_n - v_n \quad (5)$$

Finally, each block  $v_n$  is modified to be more like  $R_n$  by the following formula:

$$v'_n(x, y) = v_n(x, y) + \alpha_d d_n(x, y) \quad (6)$$

where  $\alpha_d$  is a user defined weighting factor, and  $\lambda_{dn}$  is a local weighting factor calculated using the Noise Visibility Function (NVF) to control the strength with which the autocorrelated noise  $v'_n$  is mixed with the extracted noise  $v_n$ . The local weighting factor  $\lambda_{dn}$  is a sub-block of  $\lambda_d$  that corresponds to  $v_n$ . The following formula illustrates that  $\lambda_d$  is defined for the whole image by

$$\lambda_d = \frac{1 - NVF \cdot S}{1 - NVF \cdot S_1} \quad (7)$$

In (7),  $S$  and  $S_1$  control the weight given to the autocorrelated noise based on the textured and smooth areas of the cover image. The values  $S = 3$  and  $S_1 = 1$  were used as Lee *et al.* [1] chose. The values for NVF are defined by

$$NVF(x, y) = \frac{1}{1 + \frac{D}{\sigma_{max}^2} \sigma^2(x, y)} \quad (8)$$

where  $\sigma^2(x, y)$  is the local variance of the cover image,  $\sigma_{max}^2$  is the maximum of the local variance, and  $D \in [50, 100]$ .

After  $v'_n$  has been calculated for all  $n$ , the segments  $v'_n$  together form the autocorrelated noise  $E'$ , which is added back to the filtered image  $fI$ , given by

$$I'(x, y) = fI(x, y) + E'(x, y) \quad (9)$$

## 2.2 Watermarking embedding algorithm

Once the autocorrelated image  $I'$  has been generated, its 2-level wavelet transformation is taken. The sub-bands are represented by  $I_j^\theta$ , which represents the  $\theta$ -direction sub-band in the  $j$ th level of the wavelet transformation of  $I'$ . The directions are represented by  $\theta$  as follows:

$$\theta = 1: \text{horizontal}, 2: \text{diagonal}, 3: \text{vertical}.$$

Two watermark patterns are generated. The watermarks should be of size  $M/2 \times M/2$  for the first level sub-bands, and  $M/4 \times M/4$  for the second level. The watermarks should be fractional numbers between  $[-1, 1]$ , and generated by a user key so that the watermark can be generated again for detection.

The watermarks are embedded in the image as shown by

$$I_j^{\alpha'} = I_j^\alpha + \lambda_j W_j(x, y) \quad (10)$$

where  $\alpha$  is a global weight, and  $\lambda$  is a local weight. During the embedding step,  $\lambda$  is calculated in much the same way as the noise correlation step, but it is calculated for each sub-band of the wavelet transform. The formula for  $\lambda$  is given as

$$\lambda_j(x, y) = L_j \cdot \frac{1 - NVF_j(x, y) \cdot S}{1 - NVF_j(x, y) \cdot S_1} \quad (11)$$

where  $\Theta^0$  and  $L_j$  are weighting factors that vary depending on the sub-band and wavelet transform decomposition level. The weighting factor  $\Theta^0$  is set to 2 when  $\theta = 2.5$ , otherwise it is set to 1. This gives additional weight to the watermark when being embedded in the diagonal sub-bands of the cover image, as noise is more difficult for humans to perceive in the diagonal sub-bands.  $L_j$  is a user-defined weighting factor that specifies the weight given to the watermark signal with respect to the wavelet transform decomposition level. In this experiment, we chose  $L_1 = 1$ ,  $L_2 = 2.5$ . The watermark is embedded less aggressively in the first-level sub-bands because the watermark is more easy to see when embedded in the first-level sub-bands. More weight is given to the watermark in the second-level sub-bands because they are more likely to survive a geometric or removal attack because of their increased low frequency. As before,  $S$  and  $S_1$  are once again user defined weights for the textured and smoothed regions respectively.  $S$  is set to 5, and  $S_1$  is set to 1. This insures that the

watermark is embedded more aggressively in areas with more texture.

The values for the NVF are calculated for each sub-band of each decomposition level as given by

$$NVF_j^2(x, y) = \frac{1}{1 + \frac{D}{\sigma_j^2(x, y)^2} \sigma_{jmax}^2} \quad (12)$$

where  $\sigma_j^2$  is the local variance of the given sub-band and decomposition level, and  $\sigma_{jmax}^2$  is the maximum of  $\sigma_j^2$ .

### 3. The proposed extraction scheme

It is well known that a change to the wavelet transform of an image will be localized to the corresponding area of the image in the spatial domain, so it makes sense that the watermark signal's periodicity can be extracted from the spatial domain. To do this, the Wiener filter is applied to the marked image, and the resulting filtered image  $fI$  is subtracted from the marked image  $I$  as given in (1). The resulting signal  $E$  contains the periodicity of the watermark.

Since the periodicity of the watermark can be extracted from the spatial domain, any geometric transformation on the image will be visible in the orientation of the autocorrelation peaks of the marked image.

In order to find the autocorrelation peaks, a FFT-based correlation method is used, given by

$$ACF = \frac{IFFT[FFT[E] \text{CONJ}[FFT[E]]]}{E^2} \quad (13)$$

where the operation  $CONJ(x)$  indicates taking the complex conjugate of the operand.

The resulting signal  $ACF$  will show periodic peaks which can be used to estimate and reverse a geometric attack. In order to do this, a reliable method of isolating the true peaks from surrounding noise must be used. In this experiment, a window of size  $M/2 \times M/2$  is used to find local maximums of  $ACF$ , which are taken to be possible peak candidates. Of these candidate peaks  $ACF'$ , those whose value exceeds an adaptive threshold, given by

$$ACF'(x, y) = \mu_{ACF} + \alpha_{ACF} \sigma_{ACF} \quad (14)$$

where  $\mu_{ACF}$  and  $\sigma_{ACF}$  are the mean and standard deviation of  $ACF$ , and  $\alpha_{ACF}$  is a user defined constant which varies depending on the image, and the nature of the attack.

The method proposed by Lee *et al.* [1] to choose the appropriate "base peak pair" did not work all the time in practice. The base peak pair should be the two nearest

actual peaks to the center of the image in the vertical and horizontal patterns. Using their position, a triangle is formed, and the affine transformation  $A$  is found which translates the triangle  $C$  to  $C'$ , which can be found using the period of the watermark ( $M$ ).

To consistently choose the correct peaks, every peak pair is given an initial weight prior to using Lee's method of extraction [1]. Since base peaks will more likely share the same distance from the center of the image, peaks closest to distance  $M$  from the center, while having multiple peaks with matching distance will be given a higher weight. The peak pair with the maximum weight is chosen as the base peak pair, and using  $A$ , the geometric attack is reversed.

#### 3.1. Extracting the watermark

A shift in the spatial domain does not necessarily correspond to a shift in the wavelet transform domain. To compensate for this, four wavelet transformations must be taken of the image, representing a one pixel shift in every possible direction—that is, shifted by (0, 0), (0, 1), (1, 0), (1, 1) pixels on the x, y axis. The wavelet transformation is then taken for each one of these shifted images on each level, which gives a total of 20 sub-bands representing all possible shifts in the spatial domain.

To reduce the input data for the wavelet transformation, the corrected image is split into blocks of size  $M \times M$ , and the average of these blocks  $avI$  is found. This is used as the input for the shifting of the sub-bands. The correlation between the averaged, shifted sub-bands and the watermark is then calculated as

$$corr_{j,n}(x, y) = \frac{IFFT[FFT[avg_{j,n}(x, y)] \text{CONJ}[FFT[m_j]]]}{[avg_{j,n}]^2 [m_j]^2}$$

where  $avg_n^0$  is the shifted, average block  $n$  ( $1 \leq n \leq 4$  when  $j = 1$ ,  $1 \leq n \leq 16$  when  $j = 2$ ) and  $m_j$  is the watermark signal for level  $j$ , of size  $M/2 \times M/2$  or  $M/4 \times M/4$ .

Only one shift will be correct, so one value of  $corr_{j,n}^0$  for each decomposition level  $j$  will have the highest correlation with the watermark. The highest values for each decomposition level are chosen as  $Max_j$ . These two values are then compared to two thresholds,  $T_j$ , which is calculated as follows:

$$T_j = \mu_{j_n} + \alpha_{j_n} \sigma_{j_n}$$

where  $\mu_{j_n}$  and  $\sigma_{j_n}$  are the mean and standard deviation of  $corr_{j,n}$ , and  $\alpha_{j_n}$  is a user defined constant.

If  $Max_1 > T_1$ , or  $Max_2 > T_2$ , the image is marked.

#### **4. Experimental results**

For testing, we implemented the ACF-based watermarking scheme in the spatial domain proposed by Lee *et al.* [1]. The images were processed with a combined removal and geometric attacked. Six images were used, with a total of 33 attacks conducted on each image.

The generated images were analyzed before the attacks were conducted on them, and the relative strength of the ACF peaks was recorded for both schemes. Overall, the proposed method had higher relative peak values than the previous method.